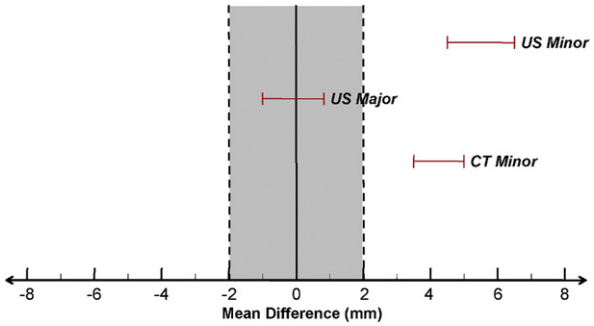

Fig 7. The 95% confidence intervals (P = .05) are shown for mean differences obtained from Bland-Altman analyses.

CONCLUSIONS

Despite these limitations, our findings illustrate the accuracy of aortic aneurysm diameter measurements in post-EVAR US surveillance, when performed by certified vascular technologists in an accredited vascular laboratory and according to a carefully devised and standardized US protocol, as described in Methods. Although the high variability suggests that one cannot use US imaging and conventional CT measurements interchangeably, consistent use of an US protocol will likely provide aortic diameter measurements that are of equal or superior reliability to the current CT measurements in post-EVAR patients. Specifically, the major axis measurement by US imaging appears to be the more accurate estimation of the aortic diameter, as reflected by the 3-D CT measurements. Further investigations on standardization of US surveillance protocol may demonstrate improved accuracy of US imaging.

AUTHOR CONTRIBUTIONS

Conception and design: FW
Analysis and interpretation: SH, KP, VR, FW
Data collection: SH, KP, SP
Writing the article: SH, FW
Critical revision of the article: SH, KP, VR, FW
Final approval of the article: FW
Statistical analysis: SH, KP, AB
Obtained funding: FW
Overall responsibility: FW

REFERENCES

1. d'Audiffret A, Desgranges P, Kobeiter H, Becquemin JP. Follow-up evaluation of endoluminally treated abdominal aortic aneurysms with duplex ultrasonography: Validation with computed tomography. J Vasc Surg 2001;33:42-50.
2. Meier GH, Parker FM, Godziachvili V, Demasi RJ, Parent FN, Gayle RG. Endotension after endovascular aneurysm repair: the Ancure experience. J Vasc Surg 2001;34:241-7.
3. Torsello GB, Klenk E, Kasprzak B, Umscheid T. Rupture of abdominal aortic aneurysm previously treated by endovascular stentgraft. J Vasc Surg 1998;28:184-7.
4. Balm R, Jacob MJ. Use of spiral computed tomographic angiography in monitoring abdominal arotic aneurysms after transfemoral endovascular repair. Tex Heart Inst J 1997;24:200-3.
5. Aarts NJ, Schurink GW, Schultze Kool LJ, Bode PJ, van Baalen JM, Hermans J, et al. Abdoimnal aortic aneurysm measurements for endovascular repair: intra- and interobserver variability of CT measurements. Eur J Vasc Endovasc Surg 1999;18:475-80.
6. Cayne NS, Veith FJ, Lipsitz EC, Ohki T, Mehta M, Gargiulo N, et al. Variability of maximal aortic aneurysm diameter measurements on CT scan: significance and methods to minimize. J Vasc Surg 2004;39:811-5.
7. Jaakkola P, Hippelainen M, Farin P, Rytkonen H, Kainulainen S, Partanen K. Interobserver variability in measuring the dimensions of the abdominal aorta: comparison of ultrasound and computed tomography. Eur J Vasc Endovasc Surg 1996;12:230-7.
8. Lederle FA, Wilson SE, Johnson GR Reinke DB, Littooy FN, Acher CW, et al. Variability in measurements of abdominal aortic aneurysms. J Vasc Surg 1995;21:945-52.
9. Sprouse LR, Meier GH, Parent FN, DeMasi RJ, Glickman MH, Barber GA. Is ultrasound more accurate than axial computed tomography for determination of maximal abdominal aortic aneurysm diameter? Eur J Vasc Endovasc Surg 2004;28:28-35.
10. Sprouse LR, Meier GH, Parent FN, DeMasi RJ, Stokes GK, LeSar CJ, et al. Is three-dimensional computed tomography reconstruction justified before endovascular aortic aneurysm repair? J Vasc Surg 2004;40:443-7.
11. Manning BJ, Kristmundsson T, Sonesson B, Resch T. Abdominal aortic aneurysm diameter: a comparison of ultrasound measurements with those from standard and three-dimensional computed tomography reconstruction. J Vasc Surg 2009;50:263-8.
12. Ashton HA, Buxton MJ, Day NE, Kim LG, Marteau TM, Scott RA, et al. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomized controlled trial. Lancet 2002;360:1531-9.
13. The U.K. Small Aneurysm Trial: design, methods and progress. The UK Small Aneurysm Trial participants. Eur J Vasc Endovasc Surg 1995;9:42-8.
14. Chaikof EL, Blankensteijn JD, Harris PL, White GH, Zarins CK, Bernhard VM, et al. Reporting standards for endovascular aortic aneurysm repair. J Vasc Surg 2002;35:1048-60.
15. McPharlin M, Rumwell C. Vascular technology: an illustrated review. Pasadena, CA: Davies Publishing; 2009.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;2:307-10.

Submitted Sep 22, 2009; accepted Jan 14, 2010.

DISCUSSION

Dr Niren Angle *(San Diego, Calif)*. Members and invited guests of the Western Vascular Society. It is my privilege to be invited to discuss a paper at this meeting and, moreover, to discuss a very interesting paper such as the one that was just presented. The USC [University of Southern California] group has put forth the proposition that ultrasound measurement of aortic aneurysm diameter is best correlated with centerline measurement of diameter on 3-D [three-dimensional] CT [computed tomography] reconstruction. In doing so, they have offered up a thesis, which at least intuitively, creates cognitive dissonance in my troubled mind but is certainly supported by the data that they present.

The authors stipulate, and I think correctly, that measurement of aneurysm diameter on axial CT is subject to much variability and error depending on the fastidiousness of the examiner and the plane in which the aorta is measured (ie, elliptical vs straight line). In this regard, centerline measurements are probably more accurate, as has been shown, but even that current belief is likely to be tempered in the future. That notwithstanding, the authors then compared ultrasound, standard axial CT, and M2S centerline measurements of aortic diameter to determine their correlation and also the variability of these measurements.

To examine this, two methods were used—linear regression and correlation, and also Bland-Altman plots. It is very useful that both these measures were used because there are limits to each method.

Correlation coefficient is a measure of the degree of association between two quantities; it does not measure how closely they agree, its use in comparing two methods that purport to measure the same thing is inappropriate, and quoting *P* values in such circumstances is meaningless. The Bland-Altman plot, which is so commonly used now in medical literature for comparative statistics, with over 10,000 citations in papers, provides a quantitative measure of how close the measurements are to each other.

The conclusion here appears to be that: (1) Using centerline major as the standard to measure by, measurement of the aortic diameter by ultrasound in the major axis is the best correlated with the least mean difference. (2) Measurement of aortic diameter in the minor axis on ultrasound had good correlation, but the mean difference in measurements was 5.38 mm in the Bland-Altman plot. (3) Axial CT measurements in the minor axis also had a mean difference of 4.25 mm.

Ultrasound is a very operator-dependent technique, and as such, it is difficult to understand that one can hope to obtain this degree of precision vs a CT scan, where although there may be variability in the measurement of the image cuts, ultrasound may have variability in the images obtained as well as the measurement. Also, the quality of the image and resolution is also dependent on many factors.

To what degree are these data only reliable at USC because of vascular technologists that have presumably honed the art of aortic ultrasound? I suspect that the interobserver variability, if the universe of ultrasound examiners was expanded beyond the ones in this study, would be much much larger than with CT scan assessment.

In reference to the Bland-Altman plots, the mean difference between CT centerline vs CT minor was 4.25 mm, whereas CT centerline vs US minor was 5.38 mm. The difference between CT centerline and ultrasound major was 0.11 mm. Does this mean that the axial CT is more accurate than is ultrasound measurement in the minor axis?

The minor axis measurement on CT is more accurate than major axis measurement and this is accepted. Why does the same principle not hold true for ultrasound major versus minor axis?

How much faith is one to have that the ultrasound technologist is reliably able to image the aorta perpendicular to blood flow, as this appears to be the requisite condition for this assessment to be valid? How does one confirm this?

In the comparison of centerline flow to ultrasound in the major axis, the mean difference between 2 measurements was 0.11 mm but the standard deviation collectively was 4.26 mm. Can you comment on this? Does the agreement between 2 measurements become less precise with different sizes of the aorta? Because if the mean difference between measurements is 0.11 mm, and this was maintained over a range of samples, why is the standard deviation so much larger, in contrast to the other comparisons?

I want to congratulate Dr Han and his colleagues and Dr Weaver for a very thought-provoking paper and for a very well written manuscript. I drank a few martinis learning the subtleties of the Bland-Altman analysis, and I feel like I know them personally. All I can say is that I am glad I am a vascular surgeon. I must say, however, unless the USC group gets a big RV to drive their ultrasound technologists all over Southern California to do these beautiful studies, it will be hard for me to be pulled away from the CT scanner as my test of choice for pre- and post-EVAR. Thank you again for the privilege and my congratulations to the authors.

Dr Sukgu M. Han. To what degree are these data only reliable at USC because of vascular technologists that have presumably honed the art of aortic ultrasound? I suspect that the interobserver variability, if the universe of ultrasound examiners was expanded beyond the ones in this study, would be much much larger than with CT scan assessment.

I think that is always a concern when it comes to any measurements that are done on US. I think if one were to pool all the US technologists in the world together, the variability would indeed be much larger, but that variability would probably come from the heterogenous nature of individual institutional protocols.

Our study and previous institutional studies by others, including Drs Sprouse, Jaakkola, and Lederle papers published in *JVS*, have shown comparable degree of variability within US measurements to axial CT scans. Therefore, this demonstrates that establishing a strict US protocol and properly training US technologists ***can*** minimize the variability all the way down to that of CT scan.

In reference to the Bland-Altman plots, the mean difference between CT centerline vs CT minor was 4.25 mm whereas CT centerline vs US minor was 5.38 mm. The difference between CT centerline and ultrasound major was 0.11 mm. Does this mean that the axial CT is more accurate than is ultrasound measurement in the minor axis?

We cannot safely say that because of the degree of variability and closeness of the mean differences. Another way to look at it is that the 95% CI [confidence interval] for axial CT and US minor measurements overlapped, so we cannot say that one is better than the other.

The minor axis measurement on CT is more accurate than major axis measurement and this is accepted. Why does the same principle not hold true for ultrasound major versus minor axis?

Well, that was the question we asked ourselves. No studies have really compared major vs minor axis measurements on US. Intuitively, we thought that if our US protocol can correct for angulation error, then we should be trusting the major axis measurements the same way that we would for the CT centerline, and our data strongly support that.

How much faith is one to have that the ultrasound technologist is reliably able to image the aorta perpendicular to blood flow, as this appears to be the requisite condition for this assessment to be valid? How does one confirm this?

After our data, we have a lot of faith in our US technologists at USC. There are two ways to ensure that this happens. First, when the aorta is being surveyed with B mode, the technologists examine the longitudinal images in order to get an idea of regions of dilatation and angulation. Second, we use color flow while the probe is being oriented. Because of pulsatile nature of aortic flow, when the probe is oriented perpendicular to it, we see a mix of alternating red and blue, compared to when it is oriented obliquely, we see mainly one predominant color. This, particularly second maneuver, is not always done, for example in radiology department US.

In the comparison of centerline flow to ultrasound in the major axis, the mean difference between 2 measurements was 0.11 mm, but the standard deviation collectively was 4.26 mm. Can you comment on this? Does the agreement between 2 measurements become less precise with different sizes of the aorta? Because if the mean difference between measurements is 0.11 mm, and this was maintained over a range of samples, why

is the standard deviation so much larger, in contrast to the other comparisons?

That was our concern as well when we looked at this relatively large variability, and that is why we performed 95% CI of mean differences from our gold standard. CI demonstrates that in comparison, the agreement seen between US major and CT centerline, and the lack thereof for US minor and axial CT minor, was a statistically significant observation. To answer your second question regarding the effect of increasing aortic size on agreement, we did not perform graded Bland-Altman plot, so I don't have a precise answer for you, but looking at the plot, it appears that the difference dots are scattered evenly with increasing size of the aorta.